

테이블 탑 디스플레이 기반의 온라인 필기 숫자 인식

김의철, 김지웅, 김수형
전남대학교 전자컴퓨터공학부
e-mail: chul3731@hanmail.net

On-line Handwritten Numeral Recognition based on Table Top Display

Eui-Chul Kim, Ji-Woong Kim, Soo-Hyung Kim
Department of Electronics and Computer Engineering, Chonnam National University

요 약

테이블 탑 디스플레이는 사람에게 친숙한 상호작용의 매개체인 손을 입력장치로 이용하는 일종의 탁자형 멀티 터치스크린이라고 할 수 있다. 본 논문에서는 이러한 환경에서 손가락 제스처를 활용하여 필기 숫자를 인식하는 연구를 수행함으로써 테이블 탑 디스플레이에 적합한 필기 숫자 인식 기술을 개발하였고, 이로 인해 추후 진행될 연속 숫자 혹은 특수기호의 성공적인 인식 가능성을 확인하였다. 실험 과정은 테이블 탑 디스플레이의 표면을 통해 입력된 손가락 궤적을 잡음제거, 대표점 추출등의 전처리 과정을 거쳐 16-방향 체인코드로 변환하고, 변환된 체인코드의 학습 및 필기 숫자 인식에 확률·통계적 모델인 은닉 마르코프 모델을 이용하였다. 학습에는 총 300개 필기 숫자 데이터를 이용하였고, 인식 실험에 사용한 별도의 100개의 필기 숫자 데이터에 대해 97%의 정인식율을 보였다.

1. 서론

최근 컴퓨터를 중심으로 한 기술 개발은 컴퓨터 사용에 필수적인 입력 도구인 키보드나 마우스에서 벗어나려는 시도가 하나의 큰 흐름을 이루고 있다. 지난 2007년 5월에 공개된 마이크로소프트사의 서피스 컴퓨터(Surface Computer)나 지난해 개발된 국내의 전남대학교 CT연구소의 테이블 탑 디스플레이(Table Top Display)가 대표적인 예로, 사용자의 양손(혹은 손가락)을 이용하여 마우스나 키보드 없이 데이터를 조작하고, 무선으로 전송하는 특징을 지닌다. 뿐만 아니라 한 사람이 아닌 다중 사용자의 사용도 가능하다. 테이블 탑 디스플레이의 주요 기술은 카메라로부터 손가락의 위치와 움직임 정보를 추출하고, 추출된 정보를 바탕으로 기존에 정의된 명령과 비교하여 사용자의 명령을 수행하도록 하는 것이다.

본 논문에서는 사용자의 손가락 제스처를 통해 필기 숫자의 입력을 받아들여 사용자가 입력한 필기체 숫자를 인식하는 기술을 제안한다. 인식에는 확률·통계적 모델인 은닉 마르코프 모델(Hidden Markov Model, HMM)을 이용하였다. HMM을 이용한 필기 문자 인식의 경우 주식회사 디오펜의 DioPen, KAIST의 BongNet, Microsoft의 인식기 등이 뛰어난 성능을 보이고 있다. 그러나 이들 모두 터치스크린 상에 전자펜으로 입력하는 환경에서 동작하는 방식으로, 전자펜같은 입력 장치가 존재하지 않은 테이블 탑 디스플레이 환경에는 적합하지 않다.

따라서 본 논문을 통해 기존의 문자인식에 있어 필수적인 도구인 마우스나 전자펜을 사용하지 않고, 손가락 제스처만으로 입력이 이루어지는 테이블 탑 디스플레이 환경에 적합한 필기 숫자 인식 기술을 개발할 수 있다.

2. 시스템 개요

2.1 시스템 구성

본 논문을 통해 제안하는 시스템은 <그림 1>과 같다. 사용자가 테이블 탑 디스플레이의 표면에 손가락으로 숫자를 필기하면, 먼저 입력된 손가락 궤적을 좌표계로 변환하여 전처리 모듈로 전송한다. 전송된 좌표는 전처리 모듈을 통해 16-방향 체인코드로 변환되고, 마지막으로 변환된 체인코드를 학습 과정을 통해 생성된 필기 숫자 0~9 각각의 은닉 마르코프 모델에 입력하여 확률을 계산하고, 이 중 가장 높은 확률을 지닌 숫자 모델을 인식 결과로 출력하게 된다.

전처리 단계는 잡음제거와 대표점 추출 과정을 거쳐 손가락 궤적을 체인코드로의 변환하는 과정으로 체인코드는 16-방향 체인코드를 사용한다. HMM의 학습 과정은 각 숫자별로 테이블 탑 디스플레이를 통해 수집한 체인코드를 이용하여 숫자별 은닉 마르코프 모델을 구성하는 과정으로, EM알고리즘의 하나인 Baum-Welch 알고리즘을 이용한다. HMM의 인식 과정은 전처리 모듈을 통해 변환된 체인코드가 어느 숫자의 은닉마르코프 모델에서 나타날

확률이 높은가를 판단하는 과정으로, 각각의 숫자 모델에 전향(Forward) 알고리즘을 적용하여 가장 높은 확률을 보이는 숫자 모델을 최종 인식 결과로 출력한다.

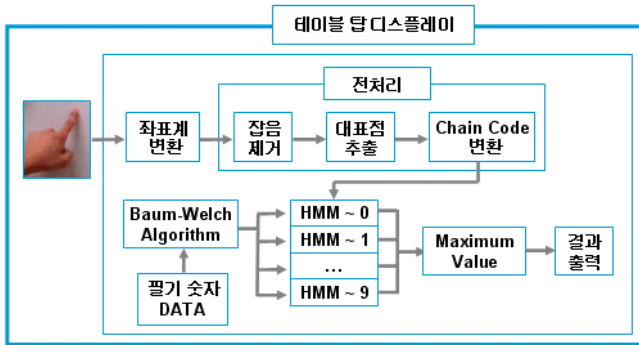


그림 1. 시스템 구성

2.2 테이블 탑 디스플레이

테이블 탑 디스플레이의 사전적 의미는 여러 가지 형태의 탁자 위에 스크린을 부착하여, 컴퓨터의 처리 결과를 눈으로 직접 확인할 수 있게 해주는 장치를 의미한다. 그러나 최근 들어 이런 사전적 의미에서 벗어나 입력 센싱 역할을 하는 표면과 출력 디스플레이 역할을 하는 표면을 서로 겹쳐놓은 직접 멀티 터치(Direct multi-touch)나 다중 사용자 테이블 탑(Multi-user table top)의 개념으로 확장되었다[1].

결국 테이블 탑 디스플레이는 <그림2>와 같이 사람에게 친숙한 상호작용의 매개체인 손을 입력장치로 이용하는 일종의 탁자형 멀티 터치스크린으로, 크게 입력신호를 감지하는 센서표면과 출력신호를 재현하는 디스플레이로 구성된다[2].



그림 2. 테이블 탑 디스플레이

2.3 은닉 마르코프 모델

은닉 마르코프 모델은 1960년대 말에 Baum의 논문에서 주식 시장 같은 여러 가지 예측 문제들의 해결에 적용되면서 시작되었다. 1970년대에 들어와 IBM의 Jerinek과 CMU의 Baker에 의하여 음성 처리 분야에서 적용되어 뛰어난 성능을 발휘하였고, 현재 문자 인식을 비롯한 패턴인식분야나, 계량 경제학 같은 응용 분야에서도 활용되고 있다[4]. 은닉 마르코프 모델은 a 행렬, b 행렬, Π 벡터로 구성되는데, 이를 은닉 마르코프 모델의 파라미터(λ)라고 한다. a 행렬은 상태(state)들 간의 전이 확률을 나타낸 행렬이고, b 행렬은 각 상태에서 관측된 심벌들의 확률 분

포이며, Π 는 각 상태들의 초기 분포를 의미한다.

2.4 전처리

사용자의 손가락 궤적이 좌표계로 변환되면, 잡음 제거(Noise Reduction), 대표점 추출 등의 전처리 과정을 거쳐, 16-방향이나 32-방향 체인코드(Chain Code)로 표현해 내게 된다. 이때 손가락 궤적이 테이블 탑 디스플레이의 표면에 그려지는 동안 외부의 빛이나 먼지등의 영향으로 인해 의도하지 않은 잡음이 포함되게 되는데, 이를 제거하는 과정이 잡음 제거 과정이다.

잡음 제거를 위해 본 논문에서는 손가락 궤적이 순차적으로 좌표계로 변환되는 특징을 이용하여, 현재 변환된 좌표와 가장 가까운 시간에 변환된 좌표의 거리를 유클리디언 거리 계산법을 통해 계산한 후, 좌표들 간의 평균 거리보다 2배 이상 먼 좌표점을 제거하는 방식을 이용한다.

테이블 탑 디스플레이의 표면에 손가락을 터치하면, 1초에 약 30개의 좌표가 변환된다. 이 좌표들을 모두 체인코드로 변환하게 되면 동일한 체인코드가 수십 개씩 만들어 지는데, 이는 메모리의 낭비와 학습 및 인식 과정에서의 언더플로우 문제를 일으키는 원인이 되므로, 필요한 좌표점만 남기고 나머지는 제거해주어야 한다. 이 과정이 대표점 추출과정으로 <그림 3>과 같이 유클리디언 거리 계산법을 이용하여 일정한 거리이상의 좌표점만 남기고, 지나치게 가까운 좌표점은 모두 제거한다.

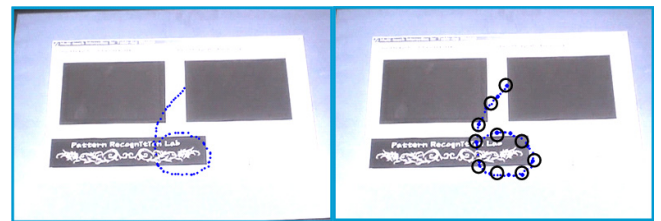


그림 3. 대표점 추출 전과 후

대표점이 추출되면 각각의 대표점과 가장 가까운 대표점 사이를 직선으로 긋고, <그림 4>의 16-방향 체인코드와 비교하여 가장 유사한 기울기를 지니는 체인코드로 변환하게 된다.

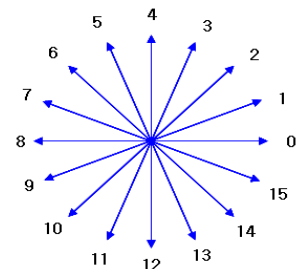


그림 4. 16-방향 체인코드

2.5 HMM의 학습 과정

은닉 마르코프 모델을 이용한 방법의 학습은 은닉 마르

코프 모델의 파라미터(λ)를 추정하는 것으로 정의할 수 있다. 이 문제는 은닉 마르코프 모델을 이용하는 데 있어 가장 어렵고 중요한 문제로, EM알고리즘으로 구현된 Baum - Welch 재추정 알고리즘으로 해결한다[4]. 학습 과정에서는 <그림 5>의 E-단계와, <그림 6>의 M-단계를 추정된 은닉 마르코프모델의 파라미터가 변화하지 않을 때까지 반복한다.

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \zeta_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (\text{식 1})$$

$$\bar{b}_j(V_k) = \frac{\sum_{t=1}^T \gamma_t(j)}{\sum_{s.t. o_t = v_k} \gamma_t(j)} \quad (\text{식 2})$$

$$\bar{\pi}_i = \gamma_1(i) \quad (\text{식 3})$$

그림 5. E - 단계

E-단계에서 <식 1>의 $\zeta_t(i, j)$ 는 은닉 마르코프 모델과 사용자의 입력이 전처리를 통해 변환된 체인코드가 주어졌을 때, 시간 t 에서의 상태 i 에서 시간 $t+1$ 에서의 상태 j 로 전이 할 확률이다. 이때 시간 t 에서의 상태 i 의 확률 $\gamma_t(i)$ 는 상태 i 로부터 j 로 전이하는 모든 $\zeta_t(i, j)$ 를 합하는 것으로 구할 수 있다. <식 2>에서 시간 t 에서의 상태 j 에서 학습 데이터로부터의 관측된 심볼들의 개수인 $\gamma_t(j)$ 를 모두 합한 것과, 시간 $t=1$ 에서 부터 $t=T$ 까지 상태 j 에서 심볼 v_k 가 관측된 개수인 $\sum_{s.t. o_t = v_k} \gamma_t(j)$ 를 이용하여 각 상태에서 각각의 심볼들이 관측될 확률을 추정할 수 있다.

$$\alpha_t(i) = P(O_{t=1}, \dots, O_{t=T} = S_i | \lambda), \text{ for } \forall t, i \quad (\text{식 4})$$

$$\beta_t(i) = P(O_{t=T}, \dots, O_{t=1} | a_1 = S_i, \lambda), \text{ for } \forall t, i \quad (\text{식 5})$$

$$\begin{aligned} \gamma_t(i, j) &= P[O_{t-1} = S_i \cdot a_t = S_j] \\ &= \alpha_{t-1} a_{ij} b_{ij}(O_t) \beta_t(j) \\ & \quad (1 \leq t \leq T) \end{aligned} \quad (\text{식 6})$$

그림 6. M - 단계

M-단계에서는 전향 변수 $\alpha_t(i)$ 와 후향 변수 $\beta_t(i)$ 를 먼저 정의하게 된다. $\alpha_t(i)$ 는 전향(Forward) 알고리즘을 통

해 구할 수 있는데, 전향 변수는 모델 λ 가 주어졌을 때, 시간 $t=1$ 에서 시간 T 까지의 각 상태(S_i)에서 전처리를 통해 생성된 체인코드가 나타날 확률을 의미한다. 한편 후향 변수 $\beta_t(i)$ 는 후향(Backward) 알고리즘을 통해 구할 수 있다. 후향 알고리즘은 전향 알고리즘의 역순으로 진행되는데, $\beta_t(i)$ 는 모델 λ 와 시간 T 에서의 상태 S_i 가 주어졌을 때, T 에서부터 시간 $t=1$ 까지 체인코드가 관측될 확률로 나타낸다.

2.6 HMM의 인식 과정

학습과정을 통해 0~9의 숫자에 대해 각각의 은닉 마르코프 모델의 파라미터(λ)가 만들어지게 된다. 인식은 관측된 체인코드가 발생할 가능성이 가장 높은 숫자의 은닉 마르코프 모델을 찾는 것이다. 다시 말하면 입력된 글자의 체인코드가 각각의 숫자 모델로부터 발생할 확률을 계산하고 가장 높은 확률을 지니는 모델을 찾는 것이 인식 과정이라는 것이다. 확률계산을 위해서 각각의 모델에 전향 알고리즘을 적용하였다. 전향 알고리즘은 전향 변수 $\alpha_t(i)$ 를 구하는 과정으로 <그림 7>과 같다. 전향 변수 $\alpha_t(i)$ 는 모델 λ 가 주어졌을 때, 시간 t 에서 관측된 체인코드 O_1, O_2, \dots, O_T 와 상태 i 의 확률을 뜻한다.

1. 초기화 단계

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N.$$

2. 유도 단계

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1, 1 \leq j \leq N.$$

3. 확률 계산 단계

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i).$$

그림 7. 전향(Forward) 알고리즘

3. 실험 환경 및 실험 결과

실험은 테이블 탑 디스플레이에서 이루어 졌다. 은닉 마르코프 모델의 훈련 데이터는 테이블 탑 디스플레이에서 0부터 9까지를 숫자를 30번씩 필기하여 총 300개의 체인코드를 만들어 데이터베이스화하였고, 이 데이터를 이용하여 학습을 수행하였다. 이 때 은닉 마르코프 모델에서의 상태(state)수는 10개로 하였고, 체인코드는 16-방향 체인코드를 사용하였다. 인식 실험은 1명의 사용자가 테이블 탑 디스플레이에 손가락으로 숫자별로 10회씩, 총 100회 필기하여 정인식율을 계산하였다.

실험 결과는 <표 1>과 같이 97%의 정인식율을 보였다. 오인식의 경우는 '0'을 필기하였는데 '3'으로 인식하였거나, '1'을 필기 하였는데 '5'나 '8'로 인식한 경우였다. 이 중 '1'

을 필기하였을 때 나타난 오인식의 경우, 테이블 탑 디스플레이의 변환 좌표는 테이블 탑 디스플레이 표면에 닿는 손가락 면적의 중간점을 의미하는데, 필기 도중 표면에 닿는 손가락의 면적은 수시로 변화하므로, 손가락을 일직선으로 움직였으나 변환된 좌표는 일직선상이 아닌 사선으로 나타난 것이 원인으로 파악된다.

표 1. 실험 결과

| 숫자 | 실험데이터 개수 | 정인식 개수 | 정확율(%) |
|----|----------|--------|--------|
| 0 | 10 | 9 | 90% |
| 1 | 10 | 8 | 80% |
| 2 | 10 | 10 | 100% |
| 3 | 10 | 10 | 100% |
| 4 | 10 | 10 | 100% |
| 5 | 10 | 10 | 100% |
| 6 | 10 | 10 | 100% |
| 7 | 10 | 10 | 100% |
| 8 | 10 | 10 | 100% |
| 9 | 10 | 10 | 100% |
| 전체 | 100 | 97 | 97% |

4. 결론 및 향후연구

본 논문에서는 테이블 탑 디스플레이 환경에서 숫자를 필기하였을 때, 실시간으로 사용자가 필기한 숫자를 인식하는 시스템을 구현하였다. 학습은 숫자별로 30개씩 총 300개의 데이터로 수행하였고, 100개의 데이터로 실험한 결과 97% 정확률을 보였다. 이는 추후 진행 될 테이블 탑 디스플레이 환경에서의 연속 숫자·특수기호의 인식의 기반이 될 수 있을 것이다.

그러나 테이블 탑 디스플레이 환경에서의 필기 궤적 샘플의 수는 아직 많이 부족한 수준으로, 다양한 사용자의 필기 궤적 샘플의 수집이 지속적으로 이루어져야만 한다. 뿐만 아니라 본 연구에서는 은닉 마르코프 모델의 훈련과정에서 상태를 명확하게 정의하지 않고 랜덤(random)으로 각각의 상태를 할당하였기 때문에, 학습을 수행할 때마다 정인식율에 있어 2~3%가량의 오차가 생겼다. 따라서 테이블 탑 디스플레이 환경에 적합한 은닉 마르코프 모델의 상태를 정의하는 연구도 병행되어야 할 것이다.

또한 은닉 마르코프 모델뿐만 아니라, 문자 인식에 흔히 사용하는 신경망등의 알고리즘과 성능을 비교하는 연구를 통해 테이블 탑 디스플레이 환경에서 문자 인식에 가장 적합한 알고리즘을 찾는 연구도 함께 이루어져야 할 것이다.

Acknowledgement

본 연구는 문화관광부 및 한국문화콘텐츠진흥원이 주관하는 전남대학교 문화 콘텐츠 기술 연구소 육성사업의 연구비 지원에 의해 수행되었음

참고문헌

- [1] 김송국, 이철우, "멀티터치를 위한 테이블-탑 디스플레이 기술 동향", 한국콘텐츠학회 논문지 제 7권 제 2호, pp. 84~91, 2007.
- [2] 송대현, 장한별, 김장운, 이철우, "테이블 탑 디스플레이 기반 지능형 증강현실 요리 콘텐츠의 개발", 한국콘텐츠학회 춘계 종합학술대회 논문집 제 5권 제 1호, pp. 34~38, 2007.
- [3] 신봉기, 김진형, "은닉 마르코프 모델을 이용한 온라인 한글 인식", 한국통신학회 워크샵 제 1권, pp. 189~194, 1993.
- [4] 한학용, '패턴인식개론', pp. 467~499, 한빛미디어, 2005.
- [5] Lawrence R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", PROCEEDINGS OF THE IEEE, VOL. 77, NO. 2, pp. 267~295, 1989.
- [6] Richard O. Duda, Peter E. Hart, David G. Stork, 유현중(ed.), '패턴인식', 2nd ed., ITC, pp. 148~160, 2006.